Towards a Theoretical Framework to Explain Root Causes of Errors in Manually Acquired Data

(Research-in-Progress)

TOM HAEGEMANS, MONIQUE SNOECK, WILFRIED LEMAHIEU, KU Leuven
FRANK STUMPE and ANTOON GODERIS, KBC Group NV

The aim of this paper is to investigate how organisations can improve the quality of their manually acquired data. This is achieved by adopting a grounded theory approach to analyse findings from an exploratory case study. The study concludes that organisations can improve the quality of their manually acquired data by increasing the intention of the data producers to input data of good quality and/or by improving the task-technology fit. This work contributes to the literature in two ways. First, we refine the theory of planned behaviour so that it can serve as a basis for initiatives to improve the quality of manually acquired data. Second, our theoretical framework demonstrates that the task-technology fit construct can not only be used to explain how the errors in the output of an information system affect the performance of a task for which this output is used, but is also relevant to explain some of the root causes of errors in the input of an information system.

Categories and Subject Descriptors: [**Methods, Concepts, and Tools for Information Quality**]: IQ Concepts, Metrics, Measures, and Models

General Terms: data quality, data entry, data producers, theory, framework

1. INTRODUCTION

The first step in the data life cycle is data acquisition [Levitin and Redman 1993, p. 217]. Despite technological advancements that allow for automatic data acquisition (e.g. RFID) or standards that enable data sharing without manual intervention (e.g. XBRL), many of the data that resides in an information system is still manually captured by a data producer [Maydanchik 2007, p. 11]. Not surprisingly, the way in which this manually entered data is acquired is an important determinant of the data quality in an organisation [Olson 2003, p. 44]. For example, in financial institutions, "manual data entry processes are predominantly confirmed to be a major data quality problem source" [Moges et al. 2013, p. 54].

To improve the quality of their manually acquired data in the most efficient way, organisations should eliminate the root causes of errors that are made during this manual data acquisition process. Correspondingly, the goal of this research is to propose a theoretical framework that can help organisations to identify the causes of errors in their manually entered data and therefore can serve as a basis to guide these future improvement actions.

In this study, we present the CEMAD (Causes of Errors in Manually Acquired Data) framework. The proposed framework is grounded in empirical observations from an exploratory case study and informed by relevant scientific theories.

The remainder of this manuscript is structured as follows: Section 2 contains the details on the adopted methodology and case study context. Section 3 presents the results of the exploratory case study. Section 4 describes which theories can be used to explain root causes of errors in manual data entry and how these theories can help to identify future data quality improvement actions. Section 5 contains the analysis of the case study results through these related theories. In addition, in this section it is explained why these theories are not perfectly suited to explain errors in manual data acquisition so as to serve as a basis for future improvement actions. Section 6 introduces the CEMAD framework, which is grounded in the empirical observations from the case study and informed by two other theories that can be used to identify root causes of errors in manually acquired data. Section 7 discusses the implications of the proposed framework for theory and practice. Section 8 contains the conclusion and directions for future research.

2. METHODOLOGY

We adopted a grounded theory approach to analyse the observations during an exploratory case study in a major Belgian financial institution. Grounded theory aims to discover theory from data [Glaser and Strauss 1967, p. 1] and has two key characteristics [Eisenhardt 1989, p. 534]: the construction of a theory should start with the collection of data before any literature is reviewed and the theory should be developed by constant comparison to the data. An exploratory case study is a case study that starts with data collection before the specification of research questions or literature review [Yin 2014, p. 238] and is therefore particularly suited to discover theory [Yin 2012, p. 29] in combination with a grounded theory approach [Eisenhardt 1989; Fernández 2004]. Accordingly, two steps were taken to build the CEMAD framework.

First, an exploratory case study was performed with the aim to investigate the root causes of errors in manual data entry that were made during the manual acquisition of home loan data. Two important home loan attributes were examined: the net monthly income of the obligors and the value of the purchased asset. The data producers who collect the data of the home loans are instructed to correctly enter these values in the home loan information system and are obliged to base these values on one or more exhibits. A possible exhibit for the value of a purchased asset can be the deed of the asset and the income of an obligor can be proven by his or her payslip. These exhibits are then sent to the main office, where they are stored in the paper archive for future consultation.

The data of the exploratory case study was collected from multiple sources:

- The exhibits in the paper archive were used to verify the correctness of the income and collateral attributes of 94 home loans.
- The documentation of the home loan information system was consulted to further understand the data acquisition process.
- Several interviews with the home loan product manager were conducted to confirm our findings and to discuss the probable causes of the identified errors.
- Four individual open interviews were conducted with four data producers in a local branch.
- The data acquisition process was observed during the entry of one home loan.

Next, we analysed the observations of the exploratory case study by comparing them against the scientific literature.

3. CASE STUDY RESULTS

We observed different small and large errors for the two loan attributes. Most errors in the income of the obligors can be labelled according three categories: *(1)* errors because of an imprecise registration (e.g. rounding errors), *(2)* errors that were made because of data entry difficulties and *(3)* an error that was the result of a situation where a data producer wanted to influence the credit score of a customer. For a small part of the errors in the income of the obligors, we were not able to directly identify their root cause.

First, some errors were caused by a number of difficulties the data producers encountered during the registration of this attribute. To start, we noticed that the data producers often had to perform complex calculations before they could enter the data of the income of the obligors. For instance, if an obligor is self-employed, the clerks need to manually calculate the data that should be entered in the database, based on the print out of a document they have to use as a guide. According to guidelines in this document, the clerks first need to interpret the tax registration forms of the obligor to obtain the net result of his/her business, the deprecations, the interests rate on investments, the exceptional charges, and the exceptional revenues. These parameters should then be used to manually calculate the tax rate which eventually should be used to obtain the current net profit. This current net profit should be manually divided by 12 and only the result should be entered in the information system. The data used as input parameters and the detailed calculation are not input in the system, but are handwritten on the paper document and stored in the paper archive. After consulting the documentation of the home loan system, it became clear that, in many cases, due to these difficulties, entering correct data was even impossible. For the income of the obligors attribute, the information system is only able to capture one value as the net monthly income, while in reality, the obligors seldomly earned the same net monthly income each month. For example, the wage of some obligors consisted of a lump sum and a part that was based on the performance of the obligor. The registered value can therefore only be an approximation of the real value.

Despite these difficulties, the data producers were very much aware of the importance of having a correct registration of the income of the obligors. For example, during the verification of this attribute, we found many pieces of paper that were attached to the exhibits with extensive calculations. In addition, while interviewing the four data producers, we were often told that they would regularly do the effort to check the results of their manual calculations on the forms with other domain experts. Furthermore, even though the aims of the study were clearly stated to the observed person, we witnessed an erroneous registration during our observation of the entry of a home loan. During this erroneous registration, the data producer spent almost 20 minutes on calculating the net monthly income of the obligor because the obligor received a weekly wage and was employed on a temporal basis. According to these calculations, the obligor earned more than 1700 euro each month. However, in the end the clerk entered 1600 euro in the information system because, in Belgium, employees that are temporary employed and paid on a weekly basis, receive their legally required bonus at the end of each week instead of the end of each year. So the clerk estimated that this extra bonus needed to be deduced from the calculated monthly income, while there was no clear guideline to do so.

Second, regardless of this general data quality awareness, we witnessed another series of errors in the income of the obligors resulting from rounding the exact income. In these cases it was clear that the data producer did not make the effort to enter the figure with a precision of two decimal places. The amount was always entered as a rounded number and in the majority of these cases the resulting error was very small.

Third, another type of error was a case where the data producer wanted to influence the credit score by entering incorrect data. In this case, the clerk entered a rounded number + 1 (e.g. 101 euro) where this rounded number was the same as the boundary value to get an advantageous credit score (e.g. 100 euro). However, in reality this obligor earned less than this registration (e.g. 99 euro).

Considering the value of the collateral, we found that all the identified errors had the same root cause. In Belgium, when buying a newly built apartment, the land and the actual construction of the apartment are often sold separately in the same transaction as this yields a tax advantage. Correspondingly, for a correct registration of such case, the data producers are expected to enter in the information system that the purpose of the home loan is to buy a piece of land and to acquire a certain construction. They have to specify this at the very start of the input process. Then, based on this input, further in the data entry process, the data producers are presented two data entry fields instead of one so that they can separately enter the price of the land and the price of the construction. However, at the beginning of the work flow, the fact that the loan request concerns a newly built apartment is often not yet clear: when the clerks ask the clients for which purpose they want a loan, they usually simply say that they want to buy an apartment. Often it is only when the clerk takes a look at the actual deed to enter the exact value of the collateral that s/he will notice that the loan concerns a newly built apartment. By then, it will require a substantial amount of effort to rectify this error because the data producers have to return to the very first step in the process, thereby losing already entered data. While in these cases it is perfectly possible to enter the correct value as no complex calculations are required, we nevertheless witnessed a recurring deficiency in the input.

Different than for the income of the obligor, the data producers did not exactly know for what purpose the value of the collateral is used. For example, while one home loan advisor was explaining for each step of the process, which data was required and how important it was for this data to be correct, he told us, without us asking, that, in the discussed case, “it doesn’t really matter whether the value of the collateral is 398 000 euro or 498 000 euro, the difference between the value of the mortgage compared to the value of the collateral is already high enough”.

4. RELATED THEORETICAL FRAMEWORKS

The factors that we empirically discovered during our exploratory case study are closely related to two established scientific theories: the theory of planned behaviour and the task-technology fit model.

4.1. The Theory of Planned Behaviour

The theory of planned behaviour [Ajzen 1991] is an established theory to explain human behaviour and is an extension of the theory of reasoned action. The theory, as shown in Figure 1, states that the behaviour of a human is determined by how much actual control this human has to perform the behaviour together with the intention that he or she has to perform the behaviour [Fishbein and Ajzen 2010]. This intention is determined by a combination of the attitude of the person towards performing the behaviour, how the person believes that others think about the behaviour and how much the individual perceives that he or she is in control of performing the behaviour [Ajzen 1991]. The Theory of planned behaviour has already been applied to data quality to explain the errors in manually entered data [Murphy 2009]. In this study, Murphy [2009] explains that the quality of manually entered data can be determined by the intention of a data producer to comply with the data entry procedure.

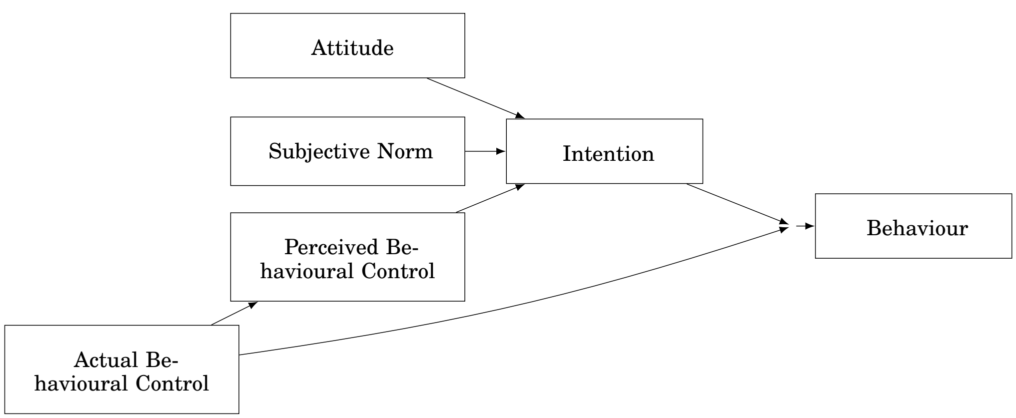

Fig. 1. The Theory of Planned Behaviour [Ajzen 1991]

4.2. The Task-Technology Fit Model

A second relevant theory is the task-technology fit model (TTF). According to the task-technology fit model, shown in Figure 2, the performance of an individual in executing a task is determined by the "degree to which a technology assists an individual in performing his or her portfolio of tasks" [Goodhue and Thompson 1995, p. 216] and by the "behaviour [of the individual] of employing the technology in completing tasks" [Goodhue and Thompson 1995, p. 218]. In other words, the performance of an individual in executing a task is determined by the task-technology fit[1] and the utilisation of the technology [Goodhue and Thompson 1995]. In practice, this model is frequently used to explain how the errors of the output of an information system affect the performance of the task for which this output is used [see e.g. Goodhue 1995, p. 1831]. To the best of our knowledge, this model has not yet been applied in a data acquisition setting, while it could: the quality of the data that the data producer entered can be considered as the performance (data quality) of an individual (data producer) in executing a (data entry) task.

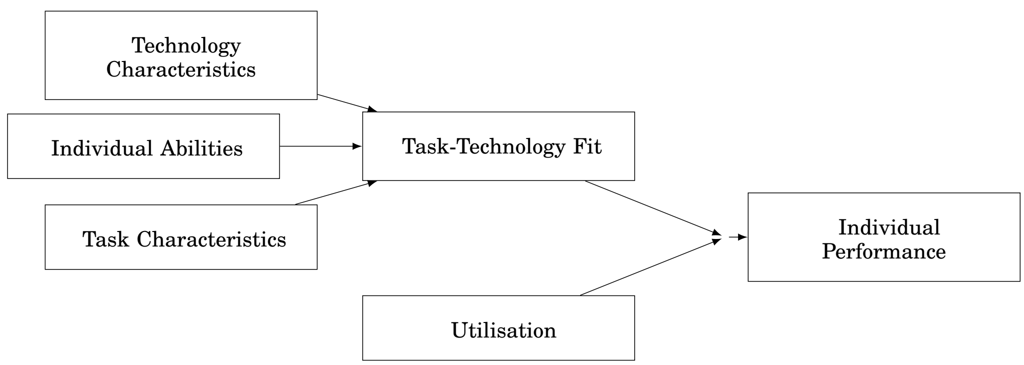

Fig. 2. A Task-Technology Fit Model [Goodhue and Thompson 1995]

5. ANALYSIS OF THE CASE STUDY RESULTS THROUGH THE RELATED THEORETICAL FRAMEWORKS

The analysis of the case study using both theoretical frameworks can be found in Table I. Due to the grounded theory approach which requires to postpone the literature review until after the empirical observations, we do not have sufficient empirical data to make statements about the subjective norm and the perceived behavioural control of the data producers. These antecedents are therefore not included in the table. Furthermore, by analysing the case study data, we came to the conclusion that, in the context of data entry, the construct of task-technology fit is conceptually the same as the degree of actual behavioural control of the theory of planned behaviour. Indeed, when an information system (technology) fits the reality (task) and the individual abilities of the data producer, it offers the data producer the right opportunities and resources to enter data of high quality and thus provides him or her with a high degree of behavioural control. In response, we merged the analysis of both constructs in Table I.

According to the theory of planned behaviour, the errors in manually acquired data are caused by the data producer's intention and/or by his or her ability to enter correct data. The model fits the witnessed errors: In the case of the income of the obligors,

[1]Goodhue [2006] has indicated that "a more accurate label for the construct might be task-individual-technology fit, but the simpler TTF label is easier to use" [p. 190]

Table I. Analysis of the case study results using the theory of planned behaviour and the task-technology fit model

Construct	Attribute: Income of an Obligor	Attribute: Value of an Asset	Is Included In: TPB	Is Included In: TTF
Attitude	In general, the clerks knew why it was important to correctly enter the income of the obligors in the information system. Even when it was difficult or impossible to correctly enter the data, the data producers performed complex calculations and verified the correctness of these calculations by consulting other colleagues. However, in some cases it was clear that the data producer did not make the effort to enter the figure with a precision of two decimal places because the registration was always a rounded figure and in the majority of the cases the errors were very small. In one case, the clerk intended to influence the credit score by entering incorrect data.	The data producers were not informed of the different purposes for which the data of the value of the asset serves. Consequently, the importance of the quality of this data is not clear.	Yes	No
Task-technology fit/actual behavioural control	The home loan application did not always support the data producers in entering the correct income of the obligors. The data entry form in the application forced the data producers to enter a single net monthly income, while in reality the obligors seldomly earned a fixed monthly income. In response, the data producers often had to perform complex calculations. Sometimes, the clerks were simply not able to enter the data correctly because they did not receive clear instructions on which calculations to perform.	The data producers were able to correctly enter the value of the collateral in the home loan application. However, in some cases, the clerks discovered some essential information to enter the data correctly at the end of the process after which the information system forced the clerks to return to step one in the process.	Yes	Yes
Utilisation	The data producers were obliged to use the home loan application to enter the data of the income of the obligors.	The data producers were obliged to use the home loan application to enter the data of the value of an asset.	No	Yes

the intention for a correct registration was high because the data producers were witnessed to have a good attitude towards entering high quality data. But, the ability to enter correct data was handicapped by a misfit between reality and provided input facilities. In the case of the value of the purchased asset, it was not really difficult to enter correct data, but the intention of the data producers seemed significantly lower because the attitude of these producers was not as positive.

Despite this high explanatory power, the construct of actual control, which accounts for the ability factor, appears to be too abstract to serve as a foundation for concrete data quality improvement actions. Actual control is defined as the degree to which a person has the required resources and opportunities to perform the behaviour [Ajzen 1991, p. 183]. Therefore, in order to propose data quality improvement actions, it is necessary to identify *which* resources and opportunities data producers require to manually enter data of high quality.

The latter problem can be addressed by considering the task-technology fit construct of the task-technology fit model: this construct provides an indication on which resources and opportunities data producers require to manually enter data of high quality. For example, in our case study, the quality of manually acquired home loan data will be increased if the information system (technology) is able to capture the real world (task). Therefore, when applied in a data entry context, the task technology fit construct is conceptually the same as the degree of actual behavioural control of the theory of planned behaviour, which suggests the addition of the task technology fit model to the theory of planned behaviour to obtain more explanatory power.

Considering the task technology fit model, we also come to the conclusion that this model on its own is only able to explain a certain part of the errors that were identified during the exploratory case study. In this model, utilisation is defined as the "behaviour [of the individual] of employing the technology in completing tasks" [Goodhue and Thompson 1995, p. 218]. The task-technology fit model therefore departs from the assumption that the individual is free to decide whether or not to use a certain technology to perform his/her tasks. However, in the case of manually acquired data in organisations, we are interested in the behaviour of the individual *while executing* tasks for which the technology *has* to be used. Even more, according to the task-technology fit model, the construct of utilisation "does not need to be considered" in an organisational data entry context where the utilisation of technology is mandatory [Goodhue 1995, p. 1830]. We therefore need to consider the *intention* of the data producers to enter high quality data while using a given technology.

Based on these considerations it can be seen that neither one of these theories is on its own perfectly suited to serve as a foundation for future improvement actions, but that the combination of both theories forms a solid basis for a framework to explain root causes of errors in manual data entry.

The next section therefore introduces the CEMAD framework based on a combination of the task-technology fit construct of the task-technology fit model and the intention construct of the theory of planned behaviour.

6. AN EMPIRICALLY GROUNDED AND THEORETICALLY INFORMED FRAMEWORK TO EXPLAIN ROOT CAUSES OF ERRORS IN MANUAL DATA ENTRY

The CEMAD framework, which is depicted in Figure 3, is empirically grounded in the results of the exploratory case study and is informed by the task-technology fit model and by the theory of planned behaviour as it includes the task-technology fit construct and the intention construct of these respective theories. In this section we will present these constructs as part of the CEMAD framework, tailor the definitions of these constructs to the manual data acquisition context and explain how these constructs can serve as a basis for future improvement actions.

Errors in Manual Data Entry. An error in manual data entry is defined as a situation where the data producer did not enter the correct registration for an attribute of a real-world object that he or she observed. These errors in manually entered data can be interpreted as the data producer having a low individual performance.

At the same time, this construct is related to data quality. That is, an error in manual data entry implies that there is a deficiency in the quality of this data [Wand and Wang 1996, p. 89]. These data deficiencies are "subject to differing interpretations depending on the particular dimension of data quality involved" [Ballou and Pazer 1985, p. 153]. Thus, an error in manually entered data can be related to several intrinsic and contextual data quality dimensions [Wang and Strong 1996, p. 22] such as data accuracy, correctness, timeliness, completeness and consistency [Ballou and Pazer 1985, p. 153; Wand and Wang 1996, p. 93].

Task-Technology Fit & Antecedents of Task-Technology Fit. Task-technology fit is defined as "the degree to which a technology assists the individual [data producer] in performing his or her [manual data acquisition task]" [Goodhue and Thompson 1995, p. 216]. A manual data acquisition task entails that the individual exerts a certain behaviour that turns an observation of an attribute of a real-world object into a registration in a database.

In the context of manual data entry, the task technology fit construct is conceptually equal to the actual control construct of the theory of planned behaviour. That is, a low degree of task-technology fit can be seen as the individual having insufficient resources to perform the desired behaviour.

This construct provides guidelines to improve the data quality in organisations. To improve their data quality, organisations can concentrate on the fit between the antecedents of the task-technology fit construct: the characteristics of the task, the characteristics of the technology and the abilities of the individual data producer. The fit between the technology and the task can be improved by making sure that the structure of the database corresponds to the structure of the real world and that there is a match between the information system and the business process in which the data is entered. The fit between the technology and the individual can be improved, for example, by enhancing the design of the user interface. The fit between the individual and the task could be improved by educating the data producers, providing clear instructions or even by providing a more pleasing working environment.

User Intention & Antecedents of User Intention. In our theoretical framework, user intention is defined as the "readiness" of a data producer to enter data without errors [Fishbein and Ajzen 2010, p. 43]. According to the theory of planned behaviour, this intention is determined by the attitude of the data producers towards entering data without errors, how the data producers perceive that others think about entering error free data and how much control the data producers think to have about entering correct data.

These antecedents can also be used by organisations as a guide for improving the quality of their manually acquired data. For example, the intention of the individual data producers could be improved by providing these producers with information about why they should enter correct data [Murphy 2009]. Lee and Strong [2003] coined the term 'why-knowledge' to describe this information.

The Impact of User Intention and Task-Technology Fit on Errors in Manually Acquired Data. Errors in manually acquired data are caused by the intention of the data producers to enter high quality data and/or the degree of task-technology fit. A higher intention of the users to produce high quality data or an increase in the degree of task-technology fit will lead to a lower error rate and smaller errors. At the same time, we expect there to be an interaction effect between the degree of task-technology fit and the intention of the data producers to enter high quality data. For example, the intention of the data producers towards entering high quality data can be reinforced by increasing the degree of task-technology fit as it will affect the perceived behavioural control of these data producers [Murphy 2009, p. 1885].

7. IMPLICATIONS FOR PRACTICE

From a practical viewpoint, the CEMAD framework can be used to guide data quality improvement actions. For example, if during data quality assessment, it was found that the errors were mostly caused by a low task-technology fit, then future actions that aim to improve the quality of the assessed data should invest in closing the gap between the task requirements, abilities of the individual data producer and the functionality of the technology. During our case study for instance, we found that there was

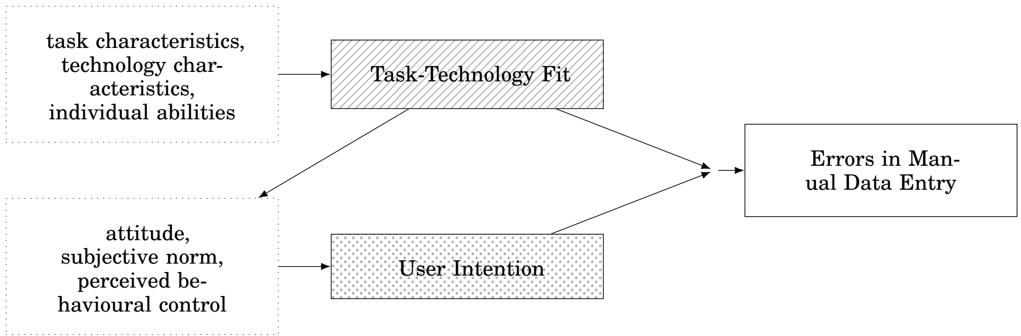

Fig. 3. The CEMAD (Causes of Errors in Manually Acquired Data) framework

a low fit between the data entry task and the technology because the information system required the data producers to enter the net monthly wage of the obligors while in reality, the obligors could receive a weekly wage, a variable wage or be self-employed. Due to this poor fit between the task and the technology, the data producers had to register the data according to various and complex business rules. As a result, the wage of the obligors was sometimes erroneously entered. To reduce the errors in the wage of the obligors, we suggested to redesign the information system so that the structure of the data would correspond to the structure of the reality.

8. CONCLUSION & FUTURE RESEARCH

The goal of this research was to propose a theoretical framework that can help organisations to explain the errors in their manually acquired data and therefore can serve as a basis to guide future data entry improvement actions.

In response, we introduced the CEMAD (Causes of Errors in Manually Acquired Data) framework. This empirically grounded and theoretically informed framework states that errors during manual data entry are caused by a poor intention of the data producers to appropriately enter the data and/or by a low degree of fit between the data entry task, the technology and the data producer.

The CEMAD framework is informed by the existing theories on user behaviour and task-technology fit but provides more insights in two ways. First, our theoretical framework refines the theory of planned behaviour with the task-technology fit construct so that it can function as a basis for concrete guidelines on data quality improvement. Second, the proposed framework demonstrates that the task-technology fit construct, which is mostly used to explain how errors in the *output* of an information system affect the performance of a task, is also relevant to explain errors in the *input* of an information system, i.e. errors in manual data acquisition.

Because our framework is informed by two existing scientific theories, we believe that it will also be applicable to explain root causes in manual data entry tasks other than the one investigated in our case study.

In future work, we aim at further validating the CEMAD framework using an experimental approach. In addition, we want to determine the economical effectiveness of the identified factors in improving the quality of manually acquired data. Furthermore, we need to investigate the impact of employing the task-technology fit construct in the context of manual data acquisition. For example, in its original context, task-technology fit is frequently measured by user evaluations [Goodhue 1995]. However, we found weak evidence that, in the context of manual data acquisition, user evaluations might not be an adequate technique to measure this construct.

ACKNOWLEDGMENTS

Part of this research was funded by KBC Group NV. The authors would like to thank Lut De Lombaerde and Iwan Barrez.

REFERENCES

Icek Ajzen. 1991. The Theory of Planned Behavior. *Orgnizational Behavior and Human Decision Processes* 50 (1991), 179–211.

Donald P Ballou and Harold L Pazer. 1985. Modeling Data and Process Quality in Multi-Input, Multi-Output Information Systems. *Management Science* 31, 2 (1985), 150–162.

Kathleen M Eisenhardt. 1989. Building Theories From Case Study Research. *Academy of Management Review* 14, 4 (1989), 532–550.

Walter D Fernández. 2004. The Grounded Theory Method and Case Study Data in IS research: Issues and Design. *Information Systems Foundations Workshop: Constructing and Criticising* 1 (2004), 43–59.

Martin Fishbein and Icek Ajzen. 2010. *Predicting and Changing Behavior: The Reasoned Action Approach*. Psychology Press.

Barney G Glaser and Anselm L Strauss. 1967. *The Discovery of Grounded Theory: Strategies for Qualitative Research* (1 ed.). Vol. 1. AldineTransaction. 271 pages.

Dale L Goodhue. 1995. Understanding User Evaluations of Information Systems. *Management Science* 41, 12 (1995), 1827–1844.

Dale L Goodhue. 2006. Task-Technology Fit: A Critical (But Often Missing!) Construct in Models of Information Systems and Performance. In *Human-Computer Interaction and Management Information Systems: Foundations*, Ping Zhang and Dennis Galletta (Eds.). M.E. Sharpe, Chapter 9, 184 – 204.

Dale L Goodhue and Ronald L Thompson. 1995. Task-Technology Fit and Individual Performance. *MIS Quarterly* 19, 2 (1995), pp. 213–236.

Yang W Lee and Diane M Strong. 2003. Knowing-Why About Data Processes and Data Quality. *Journal of Management Information Systems* 20, 3 (2003), 13 – 39.

Anany V Levitin and Thomas C Redman. 1993. A Model of the Data (Life) Cycles with Application to Quality. *Information and Software Technology* 35, 4 (1993), 217 – 223.

Arkady Maydanchik. 2007. *Data Quality Assessment*. Technics Publications.

Helen-Tadesse Moges, Karel Dejaeger, Wilfried Lemahieu, and Bart Baesens. 2013. A Multidimensional Analysis of Data Quality for Credit Risk Management: New Insights and Challenges. *Information & Management* 50, 1 (jan 2013), 43–58.

Glen D Murphy. 2009. Improving the Quality of Manually Acquired Data: Applying the Theory of Planned Behaviour to Data Quality. *Reliability Engineering and System Safety* 94, 12 (2009), 1881–1886.

Jack E Olson. 2003. *Data Quality: The Accuracy Dimension*. Morgan Kaufmann. 293 pages.

Yair Wand and Richard Y Wang. 1996. Anchoring Data Quality Dimensions in Ontological Foundations. *Commun. ACM* 39, 11 (nov 1996), 86–95.

Richard Y Wang and Diane M Strong. 1996. Beyond Accuracy : What Data Quality Means to Data Consumers. *Journal of Management Information Systems* 12, 4 (1996), 5–33.

Robert K Yin. 2012. *Applications of Case Study Research* (3 ed.). SAGE. 231 pages.

Robert K Yin. 2014. *Case Study Research - Design and Methods* (5 ed.). SAGE. 281 pages.